

Varieties of interoperability in the transformation of the health-care information infrastructure

B. A. Eckman
C. A. Bennett
J. H. Kaufman
J. W. Tenner

Health-care costs are rising dramatically. Errors in medical delivery are associated with an alarming number of preventable, often fatal adverse events. A promising strategy for reversing these trends is to modernize and transform the health-care information exchange (HIE), that is, the mobilization of health-care information electronically across organizations within a region or community. The current HIE is inefficient and error-prone; it is largely paper-based, fragmented, and therefore overly complex, often relying on antiquated IT (information technology). To address these weaknesses, projects are underway to build regional and national HIEs which provide interoperable access to a variety of data sources, by a variety of stakeholders, for a variety of purposes. In this paper we present a technologist's guide to health-care interoperability. We define the stakeholders, roles, and activities that comprise an HIE solution; we describe a spectrum of interoperability approaches and point out their advantages and disadvantages; and we look in some detail at a set of real-world scenarios, discussing the interoperability approaches that best address the needs. These examples are drawn from IBM experience with real-world HIE engagements.

INTRODUCTION

Health-care costs are rising dramatically. In 2003, the United States spent \$1.7 trillion on health care, an increase over 2002 costs at four times the rate of inflation.¹ Errors in medical delivery are associated with an alarming number of preventable, often fatal adverse events. According to a recent estimate from the Institute of Medicine of the National Academies, "... at least 44,000 and perhaps as many as 98,000 Americans die in hospitals each year as a result of medical errors. . . . Deaths due to preventable adverse events exceed the deaths attributable to motor

vehicle accidents (43,458), breast cancer (42,297), or AIDS (16,516)."²

In Australia, medication error has been estimated to result in at least 80,000 hospital admissions and a cost of at least \$350 million per year.³ The first national study in Canada of the safety of hospital

©Copyright 2007 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the Journal reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to republish any other portion of the paper must be obtained from the Editor. 0018-8670/07/\$5.00 © 2007 IBM

patients estimates that one in 13 people (7.5 percent) hospitalized has experienced an adverse event as a result of their care. Of the almost 2.5 million annual hospital admissions in Canada similar to the type studied, about 185,000 are associated with an adverse event, and close to 70,000 of these are potentially preventable.⁴

A promising strategy for reversing these trends is to modernize and transform the health-care information exchange (HIE), that is, the mobilization of health-care information electronically across organizations within a region or community.⁵ The current HIE is inefficient and error-prone. It is largely paper-based, fragmented, and therefore overly complex, often relying on antiquated IT (information technology). To address these weaknesses, projects are underway to build HIEs on the local, regional, and national levels.

On May 1, 2006, the White House stated that applying modern IT was one of the five key policies to make health care more affordable and available to all American families. The President observed that health-care providers take advantage of the most advanced technology for diagnosis and treatment, but continue to manage their medical records using antiquated paper-based filing systems.⁶ A nationwide information network will protect the privacy of a patient's medical information while making health information available in real time. We are making good progress toward the President's goal that most Americans have an electronic health record (EHR) by 2014.

Some of the HIEs being proposed are built on the local, regional, and national level throughout the world. Some regional United States examples are California,⁷ Massachusetts,⁸ Delaware,⁹ and Maine.¹⁰ Examples on the national level are Canada,¹¹ the United States,¹² and the United Kingdom.¹³ These large-scale interoperability efforts call for integrating data from a variety of organizations and agencies involved in clinical, public health, and population health information—including primary care physicians, hospitals, pharmacies, academic medical centers, and local, regional, and national public health organizations. Critical to the success of this effort is a framework to promote interoperability among health information systems, both legacy and emerging. Nevertheless, different clinical use cases are best addressed by different

interoperability approaches and architectures. For example, providing broad, integrated access to critical patient-care data by means of longitudinal EHRs might be best served by a loose federation of autonomous members of a provider network. In contrast, a population health analysis for disease management or clinical research requires periodic downloads of aggregated deidentified patient data.

This paper is a technologist's guide to interoperability. A spectrum of interoperability roles, stakeholders, and activities and major interoperability approaches are described, including a discussion of their advantages and disadvantages. We then present a look at real-world, nuanced use cases and the interoperability approaches that best fit the required needs.

STAKEHOLDERS, ROLES, AND ACTIVITIES

The roles of an interoperability solution represent the stakeholders or potential users. The following are examples:

- Providers (primary care physicians, medical specialists, emergency room physicians, hospitals, ambulatory care clinics, and long-term care facilities)
- Patients (and their families)
- Public health and biosurveillance organizations (public and private regional health agencies and national organizations such as the Centers for Disease Control and Prevention, the American Medical Association, and various public-interest groups)
- Payers (private and public, including employers)
- Clinical researchers (hospitals, pharmaceutical and biotechnology companies, academia, and government)
- Laboratories (such as pathology and radiology, public and private)

While we address payers on a general level, the focus of this paper is clinical rather than financial.

The activities of an HIE can be described as falling into two general categories: read-only and read-write. While many of these activities might also take place in a stand-alone health-care application, it is in the realm of interoperability—where data and applications are integrated across organizations, communities, and regions—that they show their

highest promise and also present the greatest information technology (IT) challenges.

The following are examples of read-only activities:

- Individual patient EHR retrieval and other activities for care delivery and management
- Data mining on aggregated data
- Alerting
- Security audits and enforcement

Example of read-write activities:

- Creating and updating an individual patient EHR
- Creating and updating privacy policy specifications

Thus, the HIE solution space may be characterized by stakeholders performing activities. Any particular HIE solution is defined by the subset of stakeholders it serves and the subset of activities it supports.

Brief scenarios of stakeholder activities are presented in the section “Read-only scenarios” and the section “Read-write scenarios” to motivate the subsequent architecture discussion and perhaps to pique the interest of computer scientists and IT professionals not yet familiar with the health-care domain. These scenarios have been gathered from health-care interoperability Requests for Proposals (RFPs) and Requests for Information (RFIs) that have recently been released, and from conversations and brainstorming with our IBM colleagues and potential customers and business partners. We provide a balance between near-term and more visionary scenarios that, together, are broadly representative of the spectrum of health-care interoperability requirements.

Read-only scenarios

Many interoperability scenarios involve retrieval of clinical information that is stored in the HIE. This section outlines these read-only scenarios, organized first by activity and then by user role.

Individual patient EHR retrieval

The fundamental unit of information of an HIE is the individual patient EHR. HIE stakeholders retrieve individual patient records for a variety of purposes. Currently this retrieval is usually a manual, paper-based, and overly complex process. In an HIE, electronic retrieval of a patient’s clinical information

by a single application—regardless of where the EHR was generated or where it is currently stored within the HIE—delivers critical information much faster and more accurately to the stakeholder who requires it.

Providers. The following scenarios involve supporting efficient and secure clinical information flow among primary care physicians (PCPs), specialists,

■ We are making good progress toward the President’s goal that most Americans have an electronic health record by 2014 ■

laboratories, and patients. In the first scenario, a PCP in private practice refers a patient complaining of dizziness, shortness of breath, and fatigue for a series of tests (blood work, chest X-ray, and cardiac ultrasound) at a variety of specialized centers. Based on the test results, the PCP reached a preliminary diagnosis of bicuspid aortic valve and referred the patient to a cardiologist at the nearest university medical center. An HIE enables electronic order submission and result reporting, delivering test results to the PCP’s desktop as soon as they are available. With the patient’s consent, the PCP allows the specialist access to the patient’s medical history and all relevant test results. The specialist reviews the clinical data, shows the patient the problem on the ultrasound, and recommends a valve replacement. The patient agrees and is scheduled for preoperative evaluation and surgery.

The next two scenarios are taken from the response of the Interoperability Consortium to the RFI related to a National Health Information Network (NHIN) issued by the U.S. Office of the National Coordinator for Health Information Technology (ONCHIT).^{14,15}

A retired military officer lives in New York City (NYC) half the year and in Florida the other half. He has chronic obstructive pulmonary disease, arthritis, and chronic pain from an old wound suffered in combat. He has identified specialists with whom he is comfortable to manage these conditions in both NYC and Florida. Some are within the United States Veterans Administration (VA) system, and some are

outside. Through the HIE sponsored by the patient's government Regional Health Information Organization (RHIO), his pain management physician sends information on his complex medication regimen to his PCP at the VA. The VA uses a decision-support tool that can detect potentially dangerous interactions among medications and can alert the patient's physicians, whether inside or outside the VA system, before orders for new medications are processed. All of the patient's providers, whether in NYC or Florida, are members of the same government HIE and have access to all of his health-care information, wherever it is generated and stored. The HIE enables the patient to keep his personal information, for example allergies, symptoms, and demographic data, up to date at all times.

A high-school student with uncontrolled diabetes mellitus is seeing an endocrinologist. His school nurse tracks his frequent blood test results and records them to be sent to his PCP and his endocrinologist. The student attempts to manage his diet with an online diary that he sends to both his physicians, along with the blood tests that he records. Both physicians have access to all this data electronically and in real time. They graph the data, identify trends and patterns in it, and share these with the student to help him understand the relationship between his diet and his condition as revealed in blood tests. This real-time data transfer is only possible because the patient, his PCP, his endocrinologist, and the school nurse are all connected electronically through an HIE.

Patients. In another scenario from the Interoperability Consortium NHIN RFI response, a PCP joins a local RHIO whose members include other physicians and the two hospitals serving the community. One of his patients chooses to join the RHIO, which enables her to access her medical records through a secure Internet site. The HIE enables her to update her personal history and payer information and schedule her next doctor's appointment online—all in one place and at a time that is convenient for her, rather than making multiple phone calls and receiving incorrect bills.

Along the lines of the high-school student with diabetes discussed earlier, a patient may want to view a longitudinal summary of measurements relevant to her particular condition in order to track her progress between visits to her physician.

Examples of measurements might be: blood sugar levels, blood pressure, thyroid-stimulating hormone levels, or measurements of lung volume. Granting patients access to their own clinical data from their home computers through the HIE gives them greater ownership of their treatment and condition, and may foster a strong alliance among physician, nurse, and patient, thus increasing patient compliance and potentially making treatment more effective.

Through an HIE, elderly or very ill patients may grant family members access to their EHR. An adult caregiver may then look up such information as her elderly father's medications or dietary recommendations to help him comply with the treatment regimen. A son may access his deceased mother's EHR to determine whether she had ever taken a particular drug, for example, Vioxx** for arthritis, and if so, the duration and dose, in order to make an informed decision about joining a class-action lawsuit against a pharmaceutical company.

Public health entities. Currently public health entities receive alerts of clinical events of potential interest, but follow-up is typically done by hand. With an HIE, both of these steps may be automated. For example, in the case of childhood injury, the injury is typically required to be reported to a public health entity. As the public health entity is alerted to the injury, an automatic process may be triggered to search through the patient's deidentified health record to look for patterns of abuse. If a pattern is detected, a qualified social worker authorized to locate the patient and investigate the incidents may be notified.

Payers. Currently, in the case of a disputed claim, providers must pull patient charts and make the charts physically available to an auditor. The charts are mailed, faxed, or hand-delivered to the auditor. This is a costly and time-consuming process that can expose patient medical information to prying eyes. In an HIE, patient charts may be made available in a restricted way to the specific auditor needing access. In addition to reducing the time between request and receipt of a chart and decreasing administrative overhead, sharing the patient records electronically reduces opportunities for medical information to be seen by unauthorized persons.

Clinical researchers. Currently clinical researchers analyze deidentified, aggregated clinical data sets to

discover significant correlations among, say, environmental exposures, treatment protocols, and clinical outcomes. But, as in the public health case, follow-up regarding individual patients whose treatment may benefit from the study results must be done manually. With an HIE, both discovery and follow-up may be automated. For example, suppose a case-control study on predictors of melanoma survival finds a strong correlation between a 10-year survival rate and a specific treatment protocol for patients whose melanomas showed a coincidence of three specific tumor characteristics (size, level of invasion, leukocyte response). Because the correlation is so strong, the researchers decide to access the EHRs of all patients with the three tumor characteristics through the HIE to obtain their name and their oncologist's name so that the oncologists may be advised of the findings and make the decision whether to enroll the patient in the treatment protocol.

Data mining over aggregated patient data

For each patient member, an HIE provides access to all clinical data generated by or currently stored at all participating provider organizations. Through participation in a RHIO, the HIE provides access to data from all the HIE RHIO members. Through participation in, say, the NHIN, the HIE may provide access to clinical data generated or stored across the entire country for each of its patient members. This extensive clinical data set offers unprecedented power to discover correlations and trends that may lead to tremendous advances in clinical care. Specialists in the field of medical informatics are only beginning to imagine the advanced analyses that will be possible. The following examples illustrate a variety of potential benefits of HIE technology, from near-term incremental improvements to more forward-looking scenarios.

Providers. A network of VA hospitals in a major city may take advantage of aggregated data downloaded from the HIE to identify patients who may be at high risk for pneumonia due to their age, medical and family history, or exposures in order to offer them pneumococcal vaccinations.

Patients. A 50-something patient in a small town who recently began synthroid therapy for hypothyroidism complains that he is noticeably anxious. His PCP lowers his synthroid dose to what he considers

the lowest therapeutic dose. When the patient continues to complain of anxiety, the PCP, who does not treat many thyroid cases, tells the patient, "You shouldn't be feeling anxiety at this dose." Because the patient is convinced that his anxiety is linked to the synthroid, he uses his HIE to mine a data set of hypothyroidism patients and discovers that 25% of the patients aged 45 to 60 who are on his dose of

■ Any particular HIE solution is defined by the subset of stakeholders it serves and the subset of activities it supports ■

synthroid report increased anxiety. He takes his findings back to his PCP. The PCP consults an endocrinologist in the closest major city, who is also an HIE member. These side effects are well-known to the endocrinologist, who recommends a decreased dose, and the patient's anxiety is alleviated while his hypothyroidism is treated effectively.

Public health entities. Through the HIE, a public-interest group interested in monitoring the public-health effects of air pollution will be able to seek and be granted access to deidentified, aggregated data from their community and surrounding communities. They may then run analyses on the data set to compare the incidence of pulmonary-related complaints in nonsmoking patients from an area whose air quality index is typically poor, with matched patients from an area whose air quality index is typically good. If significant correlations are found, the group may use this information to lobby state legislatures to enact tougher pollution control laws.

Payers. Mining a large, integrated data set representing all patients and providers in a geographical region will enable payers to identify and reward providers who deliver cost-effective, high-quality clinical care. For example, a large employer in the metropolitan New York region may compare the care provided by three large regional hospitals to employees and retirees with chronic medical conditions, such as diabetes, high cholesterol, and asthma, so that it can put pressure on the providers to improve the quality of care through a pay-for-performance plan. This might involve mining deidentified data on diagnosis, length and cost of stay, and outcome two years post-discharge. Health

maintenance organizations (HMOs) tracking expenditures may perform dimensional analyses such as *compare the length of hospital stays of patients after scoliosis surgery, by year, gender, age, and geographic region*. Payers may assess how closely providers follow National Committee for Quality Assurance guidelines in treating patients with lower-back pain, for example, by avoiding costly interventions that have not been shown to improve patient outcomes, such as MRIs or surgery, during the first six weeks of an episode.¹⁶

Clinical researchers. Data mining is a key method to identify potential subjects for clinical studies. For rare diseases and rare combinations of traits, it can be difficult to identify a sufficiently large set of matched cases and controls for a study. An HIE that spans multiple community medical centers makes this process easier. For example, researchers planning a trial of a new breast cancer drug may request a list of women with stage 3-and-above cancers who are cancer-free at least six months but no more than two years post-op.

Data mining is also critical to generating clinical research hypotheses. Again, an HIE can potentially increase the statistical power of a study by providing access to a large data set of relevant clinical data. For example, an academic research group studying skin cancer might seek to determine whether the five-year survival rates of melanoma patients who had experienced blistering sunburns before age 15 differed from the five-year survival rates of those who had not. Researchers in a pharmaceutical company might want to monitor the efficacy and safety of a newly released asthma drug by comparing a variety of adverse events in patients taking the drug with the same events in a matched control group.

Receiving alerts

In current practice, events of intense clinical interest require the sending of medical alerts. Often these alerts are generated manually and delivered in hard copy by fax. Even if automatically generated, electronic alerts often contain only the bare fact of the event, without related clinical information that can help the recipients interpret the significance of the event. The HIE not only generates electronic alerts automatically, but may include related clinical and demographic data from the patient's EHR.

Providers. With an HIE, a PCP may be notified when one of her asthma patients seeks emergency treatment at a hospital in a neighboring state due to a severe asthma episode while on a business trip. A record of all medications and treatments given is automatically sent to the patient's PCP on the patient's discharge from the emergency room.

Patients. Fred is part of a large extended family spread out over the entire North American continent. Because he rarely has contact with any but a handful of relatives living nearby, he does not know that his family has shown a greatly increased incidence of breast cancer below the age of 50 as compared with the general population. Furthermore, a gene mutation has just been discovered that tends to be present in such breast cancer families, which signals a tenfold increased chance of breast cancer in those who carry it. His HIE sends him an alert suggesting that a diagnostic test for this mutation be performed on his two 20-something daughters. One daughter is found to be negative, the other positive. The positive daughter is sent for a full pharmacogenomics workup and is put on a prophylactic drug regimen that decreases her breast cancer risk from tenfold above the average to twofold. She is also advised to have mammograms every five years. When, at age 36, she develops a small malignant tumor, it is detected in the very early stages, and her prognosis is extremely good.

In less visionary but still important examples, the HIE network sends out alerts such as the following:

- To patients on their 50th birthday to remind them to get a routine colonoscopy to detect colorectal abnormalities
- To women over 40 on a yearly basis to remind them to go for mammography
- To patients taking drugs on a long-term basis that are known to affect liver function, to remind them to go for periodic liver function blood tests

Public health entities. When occurrences of diseases of public health interest, such as tuberculosis, influenza, or anthrax, are reported, messages reporting the occurrences, along with deidentified patient demographics (e.g., sex, age, and region) and related clinical measures, may be sent to local and federal public health organizations for incorporation in their data warehouse to support further analysis. Assuming a state-wide RHIO, when the

number of occurrences of tuberculosis within a particular urban area exceeds a specified threshold, an epidemic warning may be sent to the Centers for Disease Control and Prevention (CDC) and to hospitals and providers within the area.

Payers. Currently, a physician provides treatment to a patient, and the encounter form is transferred in hard copy to the clinic's billing department. The billing department transfers the encounter information, again manually, onto the insurance company's submission form. This process is tedious, error-prone, and inefficient. In the context of an HIE, a message describing the procedure performed may be sent to the payer immediately after the patient leaves the physician's office, thus avoiding transcription errors and submission delays.

Clinical researchers. Clinical researchers who want to recruit patients for a case-control study of a new asthma drug may be alerted when a 25-year-old female patient with mild asthma and a family history of pulmonary disease is newly diagnosed or when a comparable 25-year-old patient with a family history, but without pulmonary disease, enters the clinic for a routine mammogram. (This scenario assumes that appropriate consents have been obtained.)

Laboratories. When public health agencies identify a potential outbreak concern or realize a new health risk, laboratories may be informed through the HIE to provide additional testing or look for specific indicators in test results (again assuming that appropriate consents are in place).

Security audits and enforcement

Integrating sensitive clinical data and making it available within the HIE mandates that the HIE provide state-of-the-art security and privacy features.

Providers. To demonstrate that they are compliant with national privacy regulations (such as the Health Insurance Portability and Accountability Act [HIPAA] in the United States and additional local privacy regulations) or in response to a patient complaint of breach of privacy, the HIE will enable providers, public health entities, payers, and clinical researchers to identify and report the precise data set that a specified query of a patient's EHR or

aggregated clinical data set would have returned, following the privacy policies then in force.

Patients. An HIE may provide far greater transparency to its patients with regard to accesses and uses of their medical data than current systems do. For example, suppose a patient wants to determine which parts of her medical information her employer can access. This request involves retrieving a summary page of all available information with security policy information on each element. Security information must be specified on a more granular level than the encounter level, as a patient may be willing to let her employer see that she was hospitalized for pneumonia, but not that during the hospitalization she was diagnosed HIV-positive. For more on privacy considerations, see "Privacy policy specifications" in the section "Read-write scenarios" below.

Public health entities. Public health entities are generally held to the same security and audit requirements enforced for providers. Although the public health entities in some countries allow broader access to patient information, the access must be limited to authorized individuals, and access may be logged.

Payers. Payers must comply with national and local regulations regarding the protection of patient health information. Where commercial payers are involved, protecting data is in the self-interest of the organization.

Clinical researchers. Clinical researchers are held to the same security regulations as providers. In addition, the organization with which the researcher is associated may enforce stricter rules and regulations to protect data.

Laboratories. Laboratories must adhere to the national and local laws regulating access control and access logging. Some laboratories operate by working with samples labeled only with an identifier provided by the entity submitting the sample. In this way, the patient identity is not known to the laboratory.

Read-write scenarios

Ensuring data integrity is a key function of any clinical data management system, but a large HIE faces even greater challenges. This section consists

of a brief discussion of the read-write scenarios, organized first by activity and then by user role.

Creating and updating an individual patient EHR

A patient's EHR is read and updated by many parties, with a variety of reasons for accessing it.

Providers. When a patient first enters a member organization of a RHIO, the system first checks the local, regional, and (if any) national master patient indexes (MPIs) to make sure the patient does not already exist in the system. If not, a new EHR is

■ Specialists in the field of medical informatics are only beginning to imagine the advanced analyses that will be possible ■

created for the patient, ready to be populated with individual and family medical histories, current medications, known allergies, history of hospitalizations, history of exposures, and other relevant information. The patient's minimal identity information is then registered with the regional and national MPIs. At each subsequent visit by the patient to a RHIO physician, the EHR is updated with encounter notes, diagnosis, and medications prescribed during the visit.

Patients. Currently, patients are asked to provide information on medical history, family history, exposure history, allergies, immunizations, and medications at their first visit to each provider they see. These records are updated at best during office visits, but often they are not. Patients often forget details of surgeries or environmental exposures that happened earlier in their lives. If a patient should be unconscious and need emergency treatment between visits, updated information may not be available. The HIE may enable a patient to update her EHR as soon as she is prescribed a new medication, or whenever a new allergy is noticed, or as soon as she learns that a close family member has developed cancer. The patient may also review the EHR periodically to make sure it is complete and accurate. Parents may make sure that their children's records are accurate. Patients would not be able to change information entered by doctors, at

least not without the physician's approval, but their corrections may nonetheless be recorded.

Payers. Many payers provide patients with a personal health-care record so that the patient can track his or her medical and health-related information. Payers have much of the patient's clinical diagnoses from claim submission and can build a credible medical profile from that data. Payers may also provide recommended treatment and lifestyle suggestions based on best practices for chronic illnesses.

Clinical researchers. Recall the scenario from the section "Individual patient EHR retrieval" earlier in which a case-control study on predictors of melanoma survival finds a strong correlation between 10-year survival rate and a specific treatment protocol for patients whose melanomas showed a coincidence of three specific tumor characteristics (size, level of invasion, and leukocyte response). For patients whose tumors match these characteristics, the researchers alert the oncologists to the potential efficacy of the treatment. A note may be added to these patients' EHRs that the researchers have advised the oncologists of their findings.

Laboratories. Laboratories would update the patient's EHR with laboratory reports or create a new EHR and MPI entry if the patient is not found in the MPI. Under the latter condition, an alert likely should be sent to the EHR administrator, as it may signal a false negative (e.g., the patient may actually be in the database under a different name).

Privacy policy specifications

Privacy is a complex issue, affecting and involving all HIE participants.

Providers. One way to manage HIE participation is for providers to enroll patients by default and to offer a patient global opt-out capability if they do not wish to participate. However, providers might prefer to have more granularity in their default privacy policy. For example, radiology or blood work results might, by default, be accessible to all providers in the network, but a patient's genotype or genetic profile might require a patient's explicit approval before it would be made available. Default privacy policy specifications of this kind would need to be managed by the provider.

Patients. Patients must be able to opt out of the HIE if they do not want to be involved. Patients must also have the ability to grant access to certain data and withhold it for others. For example, a patient might make his radiology or blood work results available to all physicians in the network, whereas he might prefer to grant access to his history of psychiatric hospitalization only on a need-to-know basis. Similarly, a patient might grant all network providers access to dermatopathology reports diagnostic of her dysplastic nevus syndrome, but might grant access to her full-body skin photographs only to her own dermatologist.

Public health entities. Public health entities must specify policies to enforce all national and local privacy laws. Based on the privacy laws, the circumstances under which a public health entity might be allowed to access a patient's health information may be different. In addition, those laws may affect the extent of access that a public health entity may be granted. A public health entity may be able to access only deidentified or aggregated information, in which case the health-care entity that recorded the information would be needed to correlate a patient record with the patient identity.

Payers. Payers must specify policies to enforce all national and local privacy laws. Because these policies vary widely across borders, a flexible means for defining and enforcing the policies should be sought.

Clinical researchers. Clinical researchers must specify policies to enforce all national and local privacy laws and their university's Internal Review Board guidelines.

Laboratories. Laboratories must specify policies to enforce all national and local privacy laws.

ARCHITECTURAL OPTIONS

Having surveyed a broad spectrum of stakeholders' needs and activities to illustrate the extraordinary promise of an HIE solution, we turn to a discussion of the architectural options for crafting that solution. There are four major software architectures, or interoperability approaches, available to implement an HIE: data federation, data warehousing, information distribution (one-to-many, sometimes referred to as *publish-subscribe*), and one-to-one

transactional messaging. Any specific HIE will likely be a hybrid of all four, with greater emphasis on one or the other at any given time. For example, an HIE may begin as a day-to-day clinical encounter system with a strong focus on a federation of individual patient EHRs, but evolve into a system where supporting a central data warehouse and local data marts for clinical research data mining is seen as equally important. It is critical that the HIE architecture be flexible enough to enable the relative emphasis of the four approaches to change nimbly as the needs of the HIE change.

The architectural patterns described in this section are focused on information exchange. The security architecture for authentication, authorization, and audit logging is not discussed in detail within this framework. Explicit consideration of security and privacy concerns is given as the concerns relate to specific aspects of the exchange architectural patterns.

In building our HIE offering—the IBM Interoperable Healthcare Information Infrastructure (IHII)¹⁷—we have made extensive use of the Integrating the Healthcare Enterprise (IHE) standards-based interoperability profiles, for example, the Cross-Enterprise Document Sharing (XDS) and Patient Index/Patient Demographic Query (PIX/PDQ) profiles.¹⁸ We are also active in defining the emerging SOA4HL7 standard, an activity of the HL7** (Health Level Seven**) Service-Oriented Architecture Special Interest Group.¹⁹ We are fully committed to an open-standards-based approach to interoperability. In this paper, however, the architectural patterns we describe are intended to extend beyond and, if necessary, to outlive any particular choice of standard interoperability profile. Therefore, we use general terms such as *record locator service* (RLS) and *master patient index* (MPI) instead of the IHE XDS Document Registry and PIX/PDQ Services, or the HL7 Record Location and Update Service²⁰ and Entity Identification Service.²¹

Federated (decentralized) architecture

In a federated architecture, data is distributed among a number of independent repositories. Data may be stored in multiple locations, including, for example, at multiple health-care providers, each with a multiplicity of data repositories. The central infrastructure operated by the HIE accesses this data

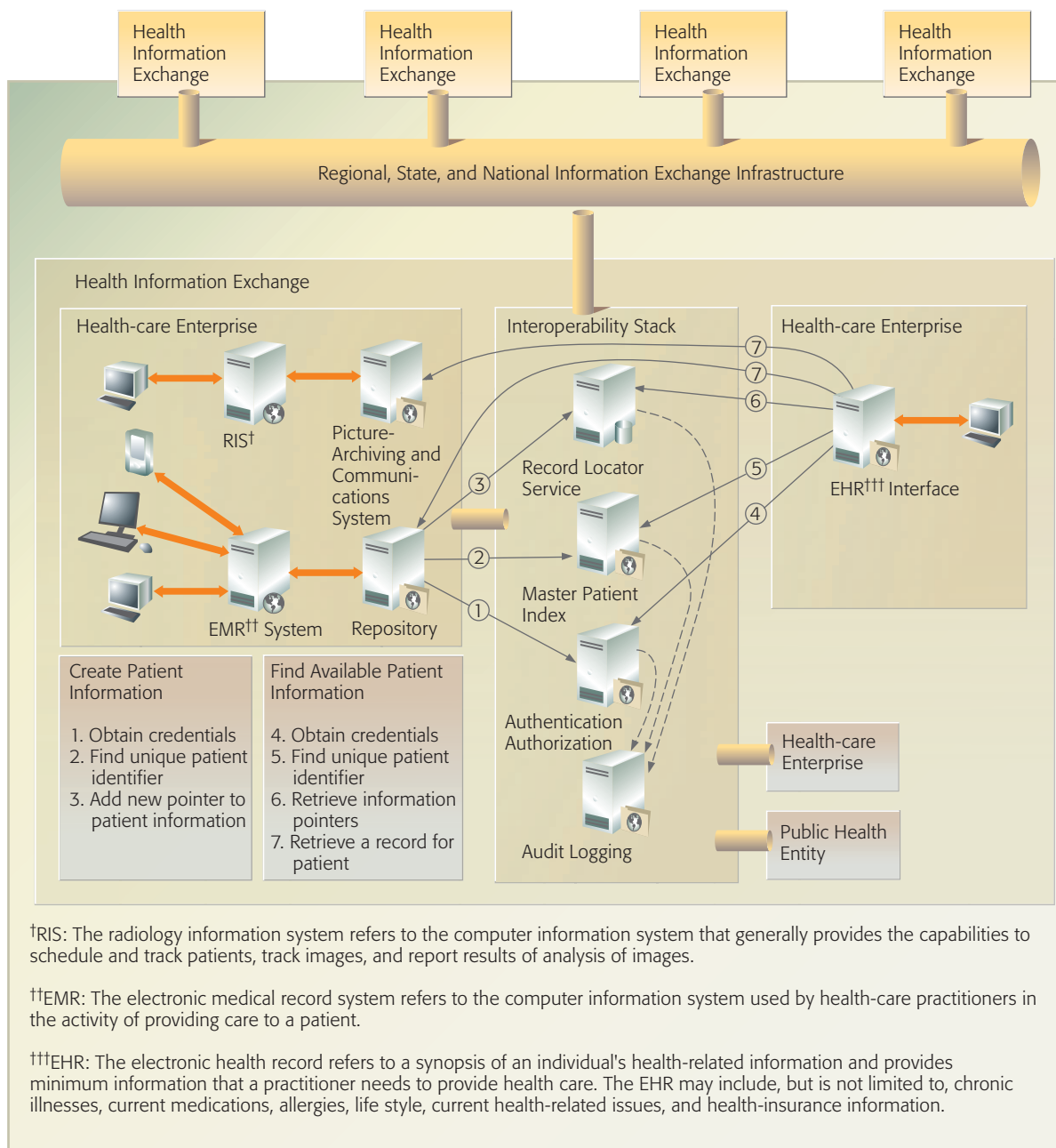


Figure 1

An example of a federated architecture; this architecture shows several independent data sources with multiple repositories at multiple provider sites

based on a central RLS and MPI. Note that the information systems within an organization may have their own independent indexes. It is the responsibility of the MPI and RLS in the interoperability stack to map to the appropriate data store.

As seen in **Figure 1**, notification of newly available patient data is sent to the central infrastructure for

accessibility. After resolving to whom the data applies, the fact that there is data available at one or more locations within the collective is logged, and metadata about the new document or documents is entered into the shared RLS and MPI. When a user wishes to access a specific piece of clinical information, a peer communication is established, and the detailed data is retrieved.

In any discussion of a federated architecture, it is important to distinguish between the distributed nature of health-care data, what one might call *de facto* federation and federation as an architectural design. De facto federation is a statement of reality. Data within a health-care enterprise is stored today in multiple distributed, autonomous systems. The federated design is an architecture intended to prevent or minimize the creation of repositories with copies of patient data. In a federated architecture, data is intentionally left in its original source location. Any health-care information architecture will have to deal with the reality of federation. The federated design is sometimes proposed as a way to increase the protection of privacy and security of patient data. A federated architecture is often considered for applications such as the longitudinal EHR, radiology image access, and laboratory results access.

Variations

The metadata held in the RLS is variable. Although a robust set of metadata at the RLS can help the user determine quickly which detailed records may be the most useful to access, there are some instances where too much metadata may reveal the nature of the clinical information, which may be more information than the user should see.

The retrieval process may involve other services provided by the interoperability hub. This could include transformation and translation capabilities to allow clinical information to be more readily interpreted by the user.

A gateway can be placed in front of one or more of the participants to allow a consistent interface for peer clinical data retrieval.

Disadvantages

The protection of privacy in a federated system with independent data sources requires implementation of security measures, privacy policies, access control, and access logging for each member of the HIE, including the interoperability stack, thereby distributing much of the burden for maintaining the functionality and reliability of the exchange to every member. In the Picture-Archiving and Communications System example, the information system being accessed by patients may well be part of a mission-critical operation in the radiology department, and it

might not make sense to expose that system to an unpredictable external workload.

Alternatively, a federated identity management solution could be implemented in which the credentials of the enterprise and a user role are used to authenticate and provide access control to clinical information. In this instance, the security and privacy are only as good as the weakest link, and a high level of trust must exist among all members of the community.

There are very significant performance exposures with a purely federated design. Response times for accessing data will vary by source and may provide a poor user experience for the physician. Also, creation of a sufficiently reliable infrastructure is difficult, as predicting points of failure and tuning the system are complex. Access to clinical information is only as reliable as the connection to the source. For example, if the original source shares data with the collective but turns off this machine for the weekend, all patient data from that source will be inaccessible until the machine is turned back on.

Advantages

In a federated architecture, the central infrastructure can be relatively small and requires less capital and ongoing operating expense. As the data does not have to be maintained outside the original source, less sophisticated arbitration is required for updating or deleting clinical information.

For certain types of clinical information, a federated architecture reduces the burden of replicating information within the interoperability stack or anywhere else within the exchange. Examples include clinical data that is accessed infrequently, such as historical laboratory results, and data sets that are very large, such as ultrasound or X-ray data.

Because detailed clinical information is shared only in a transient manner in a federation, the governance decisions regarding data sharing may be easier to negotiate; that is, the stewardship of the clinical data remains with the originating enterprise.

Warehouse (centralized) architecture

A warehouse data architecture must still deal with the reality that health-care data systems are, today, federated. However, in the warehouse design, a

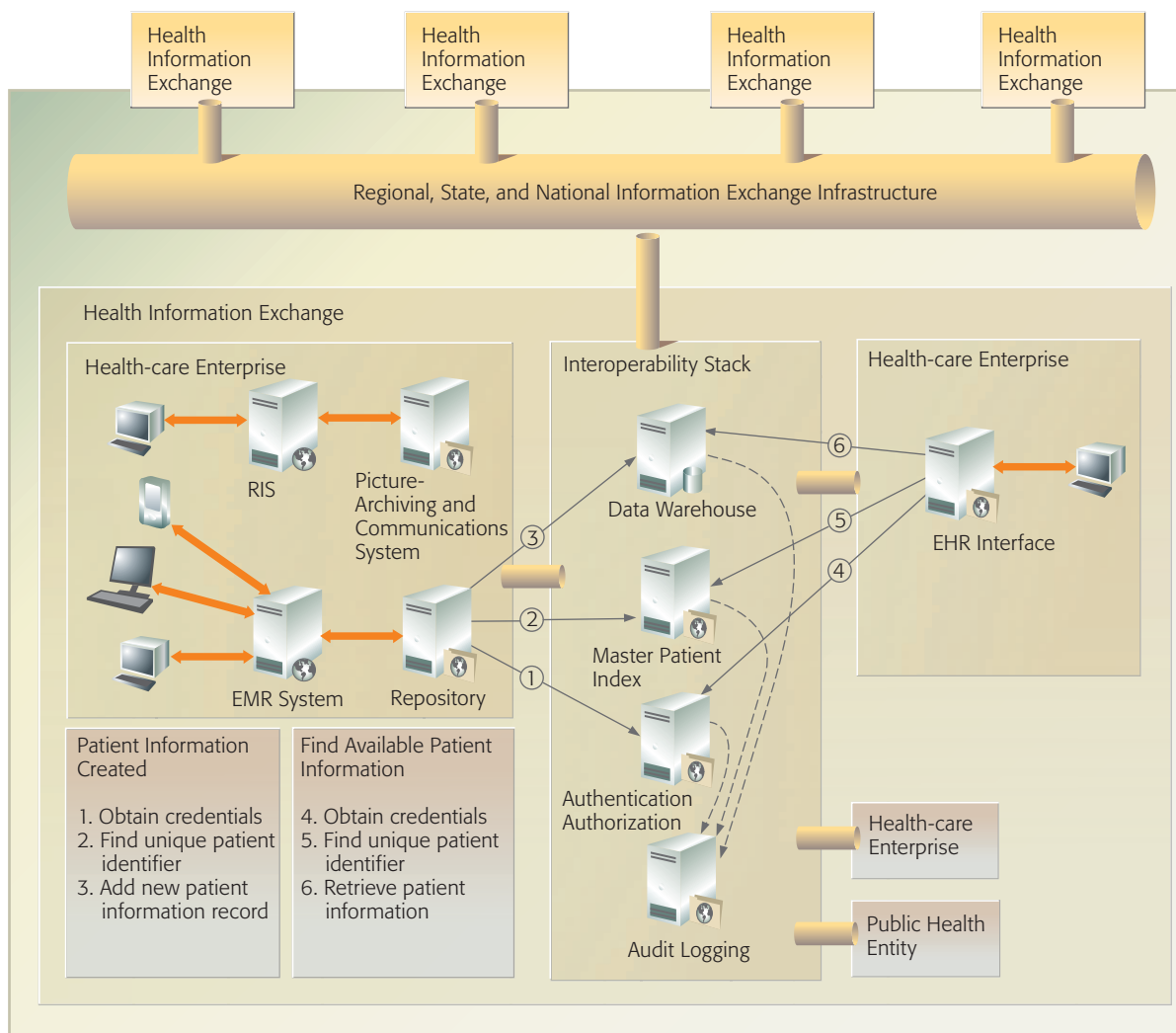


Figure 2

An example of a warehouse architecture; this architecture may involve several independent data sources; however, a data storage system, maintained in the central infrastructure, either persists or caches specific data to be shared

high-performance storage system (possibly including more than one repository for various data types) is maintained in the central infrastructure of the RHIO. **Figure 2** shows this architecture with an example interchange. A warehouse design is often considered for applications such as the longitudinal EHR, clinical trial candidate identification, disease management, population health (including warehouses of anonymized data), and clinical research (including warehouses of anonymized data).

Variations

The entire history of all patients could be stored in the central repository, which would allow for the widest variety of uses for the data.

A temporary cache of clinical information could be maintained centrally, which would limit the usefulness of the warehouse for some research purposes, but would make the most accessed information the most readily accessible for clinical care.

Clinical information that is most needed in patient care situations could be stored in the warehouse. This critical-care information would then be readily accessible and managed to allow health-care providers fast and reliable access to the information most needed in emergency and other health-care interactions. In fact, combined with a federated architecture for other clinical information that may be less critical (such as older diagnostic images and

older laboratory results), the composite architecture provides a powerful solution to many of the issues identified with the individual architectures for clinical information sharing.

Disadvantages

One disadvantage to the warehouse architecture is that it includes a more expensive central infrastructure. Although the warehouse approach does allow for easier management of reliability, response times, and overall performance, to perform these functions, the interoperability stack must provide a highly and quickly scalable deployment architecture that provides redundancy.

The interoperability stack must also be sensitive to changes in the clinical information at the source. Assume information is mistakenly entered incorrectly in a source system and is then reflected in the warehouse. If a correction is made later to the source data, a time interval exists during which another health-care provider could make a care decision based on invalid data in the central repository.

Resolving data conflicts in the interoperability stack can also be complex. Most health-care providers agree that allergy information is critical information. Because patients develop new allergies and others surcease, updates are made to a patient's allergy list. Arbitration rules must exist to define whether allergy lists are automatically replaced or merged, or whether there is a more complex workflow defined that may require human intervention to ensure consistency.

These disadvantages largely result from maintaining warehouses as materialized views of the original federated data. An alternative is to maintain a warehouse as a virtual view, using a mediator approach with a global schema that makes the heterogeneous, distributed data in the HIE appear to be part of a single, local database. IBM WebSphere* Information Integrator software, marketed as DiscoveryLink* in an earlier release,²² is an example of this technology. Although this approach ensures that only the most up-to-date data is made available through the HIE, it shares many of the disadvantages of the federated approach, including performance concerns.

Advantages

For the reasons discussed earlier, the advantages of a centralized architecture include better perfor-

mance, greater availability, and greater reliability. Arguably, a warehouse architecture can also provide improved security, privacy protection, and access control, as there is a single point of control. In a warehouse design, one can restrict the type of data stored, or even restrict the warehouse to deidentified

■ Ensuring data integrity is a key function of any clinical data management system, but the large HIE faces even greater challenges ■

versions of the data, for example, for clinical research. The particular shared data can be customized based on the access rights of the users in the RHIO, so that there would be one repository of shared data for practicing physicians, another repository of deidentified data for public health organizations, and so on.

A warehouse architecture also allows data to be cleansed and its terminology standardized by reference to a canonical controlled vocabulary or ontology. An interoperability stack connecting data sources whose terminologies have been standardized enables the consistent interpretation of clinical data across organizations within the exchange.

One-to-many architecture

In a one-to-many (or publish-subscribe) messaging architecture, each system shares clinical information that is entered into the system and processes all clinical information that it receives (*Figure 3*). The interoperability stack does not maintain a persistent store of the information, but is merely a clearinghouse for information distribution. We define the transaction hub in the interoperability stack as the component that ensures a reliable transmission infrastructure. Each published piece of data must be maintained in the transaction hub until it is delivered to all subscribers.

The interoperability stack in the one-to-many architecture model operates much like an electronic mailing list server. Each enterprise within the HIE that has data entered into its systems publishes the relevant data outward to the interoperability stack. All other enterprises participating in the HIE that have subscribed to data feeds receive the data from

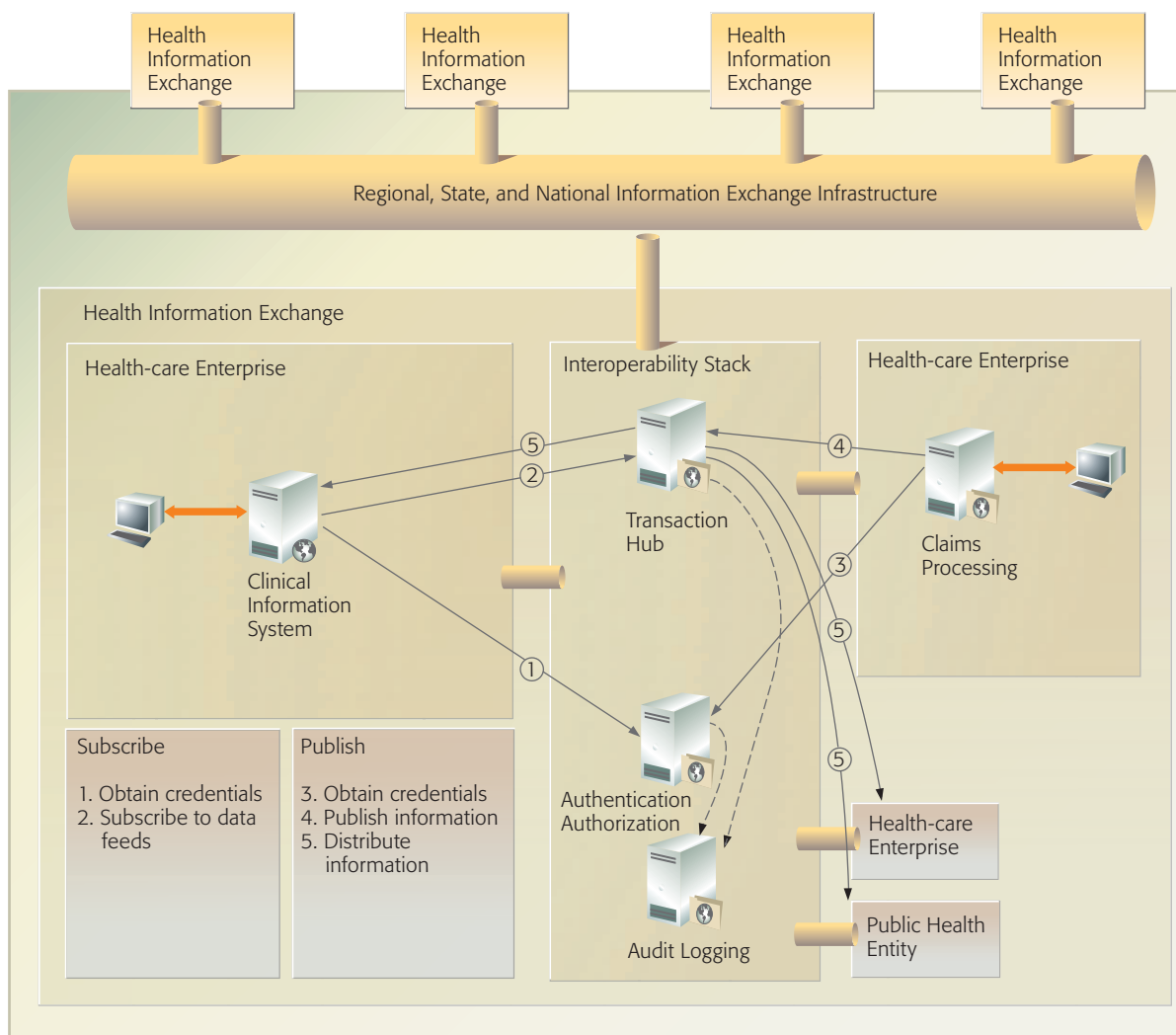


Figure 3

An example of a one-to-many messaging architecture; by allowing an organization to subscribe to information contributed to the exchange, the one-to-many model enables relevant information to be quickly disseminated throughout the exchange

the transaction hub. The receiver of the data is then responsible for processing the information that it receives.

The ultimate use of the information that a subscriber receives could be a comprehensive aggregation of all health-related data for all patients or a compilation of some subset of data received. The data received could also undergo statistical analysis processing, and only the results would be saved.

The one-to-many architecture may be used for many applications. Some examples are adverse drug reaction reporting, public health advisory distribu-

tion, epidemiology case alerts, protocol adherence, and true clinical data exchange between disparate clinical systems.

Variations

An MPI could be introduced to allow patients to be uniquely and correctly identified across enterprise boundaries. This identifier would be assigned to the patient before sending the information, and it would be mapped back locally. Identifiers must still be assigned from the central infrastructure, but it becomes the responsibility of each participating organization to provide (on sending) and understand (on receiving) the global patient identifier.

Information may be deidentified when sent out to the transaction hub. Clinical information required to be reported to public health agencies may be transmitted without identifying the individual patient associated with the clinical information.

Information may be aggregated before publishing to the transaction hub. By either combining information from multiple patients or aggregating multiple types of clinical information for a single patient, network traffic can be reduced, and a more cohesive, comprehensive snapshot of data can be disseminated.

Filters can be placed on the edges and within the hub. By filtering the clinical information to be sent only from the source or received at the destination based on meeting certain criteria, network traffic and clinical information processing can be reduced.

Disadvantages

There is usually a significant amount of tuning that must be done in order to receive just the right amount of information from the interoperability stack. Subscribing to too much information can overwhelm the data processing engine receiving data from the HIE and not subscribing to all of the information needed will result in the loss of important information.

Much information is replicated throughout the system. For most applications, excluding those that need discrete data and provide some individualized processing on each discrete piece of data, this data replication is unnecessary.

Advantages

The infrastructure for the one-to-many architecture is relatively lightweight, but it can provide a flexible basis for many types of clinical information exchange. With the message-routing capabilities needed to support this architecture, the implementation platform architecture of choice should be an Enterprise Service Bus (ESB).

The clinical information that is exchanged when using this pattern is not stored persistently within the interoperability stack. The transient nature of the information alleviates the need to negotiate information stewardship and most information access policies among the constituents.

One-to-one messaging architecture

With the one-to-one (reliable transaction) messaging architecture, each member of the exchange can communicate with other members. The interoperability stack for this pattern is the most lightweight and requires the least investment in platform components.

The architecture represented in *Figure 4* can supplant other forms of peer-to-peer communication. Frequently, the interoperability stack is used in place of e-mail, facsimile, or traditional mail delivery processes.

When implementing this pattern, a service-oriented architecture approach with an ESB should be considered. While it is possible to implement this architecture with an infrastructure that is less scalable and flexible, an exchange built on an ESB will have the foundation to support additional architectural patterns without significant investment in redesign and redeployment.

Some of the applications of a one-to-one architecture include transmission of claims and claim status requests, electronic payments, pathology laboratory results, radiology reports, and e-prescribing.

Variations

An intelligent transaction hub can be included in the interoperability stack to allow multiple atomic exchanges to be coalesced into a single transaction. By including the intelligent hub, complex transactions can be formed which, in turn, can be committed or rolled back as a single virtual transaction.

Disadvantages

Because the one-to-one pattern is the simplest pattern, the interoperability stack is usually implemented with a lightweight infrastructure that tends to need reimplementations when additional features and functionality are added to the exchange.

The HIE initiatives that this pattern supports are generally administrative or involve process automation and do not exchange subjective clinical information, such as clinical notes or treatment protocols. Although this is an advantage for near-term implementation, many of the critical information stewardship and information-sharing policies are not addressed.

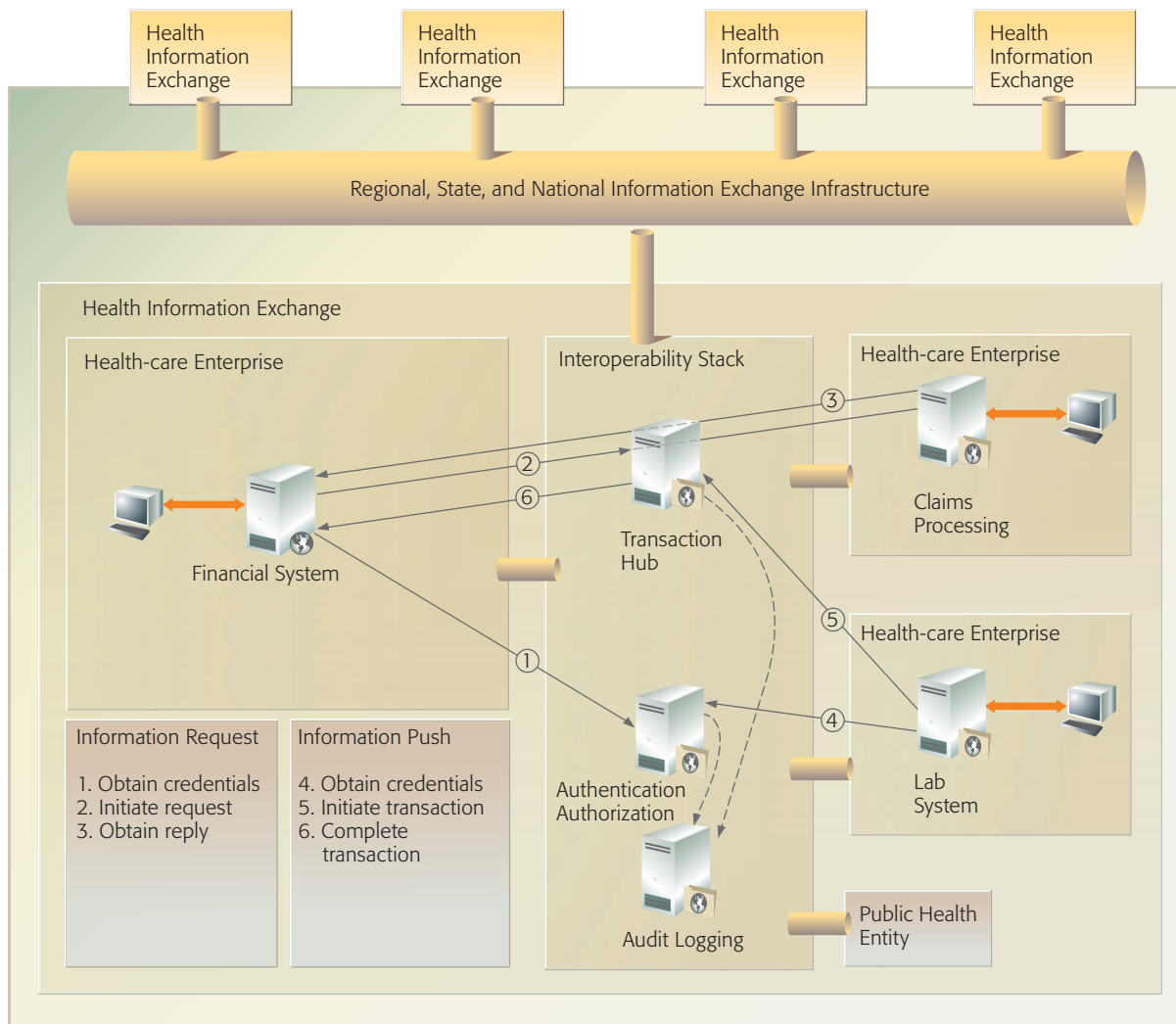


Figure 4

A one-to-one messaging architecture; this enables peers in the HIE to communicate reliably and securely with other members of the exchange

Advantages

The cost to implement the one-to-one messaging architecture can be quite low, making this a good starting point for an exchange. Starting with a system that is inexpensive to implement, maintain, and administer lets the exchange constituency focus on the harder, nontechnical issues associated with governance and collaboration, providing a business foundation for later, more sophisticated HIE initiatives.

No health information is stored within the interoperability stack, making this pattern less controversial in terms of information stewardship. Because

information is transferred point-to-point and is not persistent, there are fewer security vulnerabilities.

The one-to-one messaging architecture supports the automation of many existing processes, such as faxing or mailing reports, faxing or mailing claims, and administering claims. Because the cost to provide an existing manual process can be quantified relatively easily, and the cost of implementing the technology to support this pattern is low and also quantifiable, the return on investment for the interoperability stack is also readily quantifiable. Generally, the return on implementing an initiative on this pattern is realized in a short time.

REAL-WORLD SCENARIOS

Any real-world HIE will likely be a hybrid of all four architectures, with greater emphasis on one or the other at any given time. In this section, we describe a few key scenarios and the hybrid architectures that would be used to meet their needs. Note that applying these technologies to real-world situations will have an immediate and beneficial change on medicine as practiced today.

Administrative exchange

While the nonfunctional technical requirements involved in a clinical HIE (for example, reliability, scalability, privacy, security, and response time) are complex, the governance of the exchange and the stewardship of the clinical information that drives many of the nonfunctional requirements are even more complex. Because of these factors, it makes sense to begin with an administrative HIE.

An administrative exchange focuses on processing health-care claims. The providers and payers share an infrastructure that allows the providers to check patient eligibility for health care, provide referrals, make claims, and request claim status. Payers can send payment transactions to providers, who are able to process electronic accounts receivable. These transactions were required by HIPAA and standardized by the American National Standards Institute through the Healthcare Task Group in the X12N Insurance Subcommittee of the X12 organization.

To enable the administrative exchange, a peer-to-peer, transaction-based architecture is needed. By taking advantage of a centralized infrastructure to manage secure, reliable transactions, communities of providers and payers have realized savings in administrative overhead, errors in processing, and time to payment. As the different constituents in the community are added to the exchange, the processes that each constituent type (payer or provider) use to communicate with other constituents is also standardized on best practices, which can realize additional savings.

The administrative exchange and the community that forms around the exchange frequently begin to look beyond administrative information exchange to exchanging clinical health information. Although the administrative exchange is a convenient business organization, the lightweight transactional infrastructure needed for administrative exchange is

not well-suited to other types of exchanges. The primary constituents in a clinical exchange are also generally limited to providers, although the payers may support the clinical exchange if benefits to patient care can be demonstrated.

The Western New York HealtheNet (WNYHealth-eNet) is an organization created as a result of a

■ Any real-world HIE will likely be a hybrid of all four architectures, with greater emphasis on one or the other at any given time ■

collaboration of four health-care-provider and three health-care-payer organizations. WNYHealtheNet's charter is to share an infrastructure to optimize patient-care delivery in the western region of the state of New York. Because the infrastructure deployment costs are low and the solution is automating and optimizing an inefficient and error-prone process, the community estimates that it realizes nearly \$2.4 million in savings each year.²³

The Utah Health Information Network (UHN) is an example of an administrative exchange that is also considering clinical information exchange. The UHN is a not-for-profit entity formed by a coalition of payers, providers, and the state government of Utah. UHN forms the hub of a transaction infrastructure, with only transient data flowing through the hub between the payers and providers.²⁴

Clinical exchange for longitudinal EHRs

The most discussed scenarios center on providing clinicians with broader and faster access to patient information at the point of need. A clinical HIE is intended to provide the right health care more quickly, with reduced cost and improved outcomes.

Clinical HIEs are generally more difficult to implement. This is due to not only the lack of true interoperability standards on the technical side, but also to thorny governance and business issues on the nontechnical side. Many of these governance issues—often the most difficult obstacles to interoperability—are mentioned in the preceding sections on the various architectural patterns.

Because of the broad spectrum of functional and nonfunctional requirements placed on clinical exchanges, there are two broad categories of data architectures and two user-identity management architectures. In both categories the differences can be summarized as centralized or federated.

In North Carolina, the Western North Carolina Health Network (WNCHN) has deployed Data Link,²⁵ a system that combines a fully federated architecture with a centralized identity-management solution to create an HIE for the hospitals in the region. With Data Link, a clinician authenticates himself and, provided he has the proper access rights, is allowed to select a patient and see what clinical records are available for that patient.

In the WNCHN Data Link architecture, the only information stored within the interoperability stack is pointers to facilities. When a clinician selects a patient, each facility is queried for the metadata clinical information stored at the facility. The metadata, including the record-retrieval information, is then presented to the clinician as a longitudinal EHR.²⁶

Another approach to providing a longitudinal EHR is to provide some set of critical care information within the interoperability stack with pointers to detailed patient information maintained within the constituents. This hybrid architecture provides all of the advantages of a centralized data architecture for the most important data needed to care for an individual, while taking advantage of the benefits of the federated architecture for all other clinical information.

The United Kingdom National Health Service is developing a hybrid architecture to implement the NHS Care Records Service (CRS). At its core is the *Spine*, which provides the interoperability hub.^{27,28} The Spine stores some patient demographic information along with allergies, adverse drug reactions, and accident and emergency visits. The Spine is also responsible for providing security and interfaces to the local IT systems, which maintain the detailed clinical information.

The Danish Healthcare Data Network began as a messaging architecture and is going through a transformation to a federated architecture. When the network began in 1994, the purpose was to simply

standardize the way data was transmitted between health-care enterprises.²⁹ The realization that the Internet model of data pull could be used to provide wider access and broader functionality led to the growth of an interoperability stack that maintains access control rights and pointers to clinical data. This approach has led to the Public Health Portal, which allows Danish citizens to interact with health-care professionals.³⁰

Population health analysis

The comprehensive integrated clinical data set available through an HIE promises incredible power for disease management and clinical research—whether performed by academic researchers, hospital administrators, the clinicians of a regional independent practice association, or a pharmaceutical company. (Patient consent must be obtained before the data may be used for research purposes.) In the simplest case, ad hoc explorations of the data set can be used to generate hypotheses that can later be tested in carefully designed follow-up studies. However, if the data set is rich enough to enable closely matched populations of cases and controls to be defined, analysis of this data may indeed yield significant findings.

IBM is creating an intraoperability architecture as a solution for a large university medical system that will enable it to integrate data from its large hospitals, faculty practices, and outlying satellite hospitals and clinics. The customer prefers a federated approach for providing an essential medical data set (EMDS) for each patient, drawn from all participating providers. For clinical research, however, data from the EMDS is integrated, along with more specialized data, into data warehouses. In populating the warehouses and any specialized data marts derived from them, the data is deidentified and cleansed by using mappings between ontologies to normalize terminologies and to normalize units across data sets, and also perhaps for discretizing continuous variables. Upon analyzing the warehouse data, researchers may discover evidence that a certain subset of patients is at substantially increased risk for developing a particular disease. The researchers may then request that this subset of patients be re-identified so that they may gain access through the federated system to portions of the individual patient's EHR (specifically, their names and the names of their PCPs). The

researchers may then inform the PCPs of their findings so that follow-up tests may be performed.

IBM recently collaborated with Mayo Clinic to develop a solution for researchers wanting to combine patient demographics data from a patient registration system, diagnostic data from inpatient and outpatient billing systems, laboratory test results and reports from the primary EMR system, and clinical notes documents. At Mayo, a central repository architecture was chosen over data federation to minimize impact on the performance and availability of the source systems. Data latency requirements were considered on each data subject. For example, the nature of patient registration data made daily feeds acceptable. However, laboratory test results required near real-time feeds for scenarios such as identification of clinical trial candidates. To meet these requirements, the system used a combination of batch extract-transform-load procedures, combined with data replication and an innovative technique which favored using lower-latency sources in cases where the same data instances came from more than one source.

Clinical notes data also posed a challenge. Clinical notes contain a wealth of information; however, as they are largely unstructured, analyzing such documents throughout a patient population is time- and labor-intensive. The IBM Unstructured Information Management Architecture³¹ (UIMA) and machine-learning technology were used in conjunction with both IBM and Mayo text annotators to identify concepts within the notes and match them with a standardized and coded vocabulary. As a result, relational queries over concepts expressed in the notes could be formed easily. Compliance with security and privacy requirements was also key for this system. To meet these needs, access to the system was provided through the IBM Data Discovery and Query Builder product,³² which enabled researchers to build their queries by using business terms. The system converted the queries to Structured Query Language and, during the conversion, appended additional conditions to filter the data to meet privacy and security requirements. In addition, tight control over access to the system and auditing were implemented. This system is now being used primarily to identify patient cohorts for clinical research. However, usage and data elements of the repository are expanding to meet Mayo needs beyond research.

Population health monitoring for regulatory reporting and biosurveillance

The risk of national and worldwide disaster due to infectious disease is being increased by the rise of a global economy and the growing reliance by developed countries on global transportation and trade. This emerging trend makes the development of automated public health monitoring a high priority. The coincidence of rapidly spreading

■ Early detection of infectious disease is both essential to public health efforts and a vital first step in the fight against bioterrorism ■

infectious disease along with the rapid transportation, propagation, and dissemination of the pathogens and vectors for infection poses the risk of new and dangerous pandemics.^{2,33-35}

The anthrax attacks in the United States that occurred shortly after the events of September 11, 2001, demonstrated the subtle nature of biological attacks and their effectiveness in spreading terror.^{36,37} The effects of that particular attack were mitigated by the fact that anthrax is not easily weaponized, and the infections were limited by physical distribution of the anthrax spores (i.e., no person-to-person transmission).³⁸ Bioterrorist attacks based on other agents could be significantly more dangerous and difficult to contain.³⁶⁻³⁹

The requirements for reporting diseases are mandated by national, regional, and local laws or regulations, and the list of reportable diseases varies at each level. In the United States, the CDC provides uniform criteria for reporting cases⁴⁰ and maintains a list of nationally notifiable diseases⁴¹ considered to be significant threats to public health. Individual states also define lists of reportable diseases. As a result of this protocol, the first line of defense for developed nations against infectious disease is our world-class medical care infrastructure. In the United States, physicians depend on several important programs, many administered by the CDC, to receive early warning and prepare to respond to new epidemics. To date, these programs have proven effective in protecting society from naturally occur-

ring pathogens and food-borne illness.⁴²⁻⁴⁸ However, many current programs are optimized to help physicians diagnose or recognize new illnesses.⁴⁹ The data these programs depend upon comes, in large part, from local and regional medical institutions, laboratories, and insurance companies, such as those noted in References 42-53.

The vulnerability of our society to fast-spreading, naturally occurring agents has also been demonstrated in recent years, for example, in the SARS outbreak,⁵⁴⁻⁵⁶ cases of West Nile virus,⁵⁷ and dengue fever.⁵⁸ Early detection of infectious disease is both essential to public health efforts and a vital first step in the fight against bioterrorism. The need for an early warning system has been recognized, and a number of local and regional experimental programs are now in place.⁵⁹⁻⁶³

For example, the Real-Time Outbreak and Disease Surveillance system (RODS),⁶² an open-source project developed by the University of Pittsburgh, enables state and local governments to adopt a computer-based public health surveillance system for early detection of disease outbreaks. RODS is a syndromic surveillance method that recognizes outbreaks based on analysis of free-text chief complaints collected during triage in emergency departments and registration in acute-care clinics. Their current algorithms focus on detection of an anthrax outbreak due to outdoor aerosol release of the pathogen.

In the RODS design, hospitals send anonymized HL7 Admission, Discharge and Transfer messages containing demographic data and chief complaints by means of their internal integration engines (message routers) to a centralized RODS database. RODS automatically classifies the chief complaint portion of the text into one of seven syndrome categories by using Bayesian classifiers and then assesses whether anomalous densities of cases in space and time can be detected in the data. Anomalous findings trigger e-mail or pager alerts to infection-control personnel. RODS is built on a hybrid architecture that consists of a one-to-many messaging architecture for distributing data from the individual hospitals to the RODS hub, a centralized data-warehousing architecture for performing analyses on the free-text chief complaints, and a simple form of one-to-one messaging for alerts. RODS was used for biosurveillance at the 2002 Salt Lake City Winter Olympics and in the states of Pennsylvania and Utah.

In Missouri, the Cerner Corporation and the Kansas City Health Department (KCHD) recently collaborated on an electronic reporting system for bio-surveillance that takes advantage of a one-to-many messaging architecture and a data-warehousing architecture.⁶¹ A network of 22 laboratories extracted and published new laboratory test orders, results, and patient demographic data on a daily basis to a central data clearinghouse. There the data was loaded into a warehouse, cleansed, and normalized to a common nomenclature. Reports were then sent to the KCHD summarizing trend information on test orders and on microbiology-test results. This electronic system delivered the data to the public health department significantly earlier (by two to three days) than conventional (manual) reporting methods.

CONCLUSION

In this paper we have presented a technologist's guide to health-care interoperability. There is a broad spectrum of stakeholders, each of whom plays a different role in the future of health-care interoperability. We reviewed several important architectural approaches to interoperability and discussed their advantages and disadvantages. When designing a real-world system, one should not consider these approaches as mutually exclusive. Rather, to satisfy all of the many stakeholders in a dynamic landscape of requirements, one should build into an interoperable health-care infrastructure the ability to adapt to changing real-time requirements. This is most evident, perhaps, when considering the alternative designs for data federation against the requirements that must be met to satisfy both individual EHRs and public health. At some instant it may be desirable to keep a large part of a patient's EHR federated. At a later time (for example, before a hurricane or in response to some emerging infectious epidemic), it may be necessary to adapt to new requirements and create regional warehouses of specific data. Embracing a flexible architecture that can adapt to the changing needs of public health officials and clinicians is the best strategy to realizing a scalable HIE.

ACKNOWLEDGMENTS

The authors thank our colleagues in IBM Software Group, Sales and Distribution, Research, and Global Business Services for brainstorming with us on scenarios and architectures.

*Trademark, service mark, or registered trademark of International Business Machines Corporation in the United States, other countries, or both.

**Trademark, service mark, or registered trademark of Merck and Company, Inc. or Health Level Seven, Inc. in the United States, other countries, or both.

CITED REFERENCES

1. Health Insurance Cost: Facts on the Cost of Health Care, The National Coalition on Health Care, <http://www.nchc.org/facts/cost.shtml>.
2. L. T. Kohn, J. M. Corrigan, and M. S. Donaldson, *To Err Is Human: Building a Safer Health System*, The National Academy Press, Washington, D.C. (2000), <http://www.nap.edu/catalog/9728.html>.
3. B. H. Barraclough, "Safety and Quality in Australian Healthcare: Making Progress," *The Medical Journal of Australia* **174**, No. 12, 616–617 (2001).
4. G. R. Baker, P. G. Norton, V. Flintoft, R. Blais, A. Brown, J. Cox, E. Etchells, et al., "The Canadian Adverse Events Study: The Incidence of Adverse Events Among Hospital Patients in Canada," *Canadian Medical Association Journal* **170**, No. 11, 2004.
5. Definition and Select Characteristics of HIE Initiatives, Foundation for eHealth Initiative, <http://www.ehealthinitiative.org/pressrelease825A.msp>.
6. G. W. Bush, "President Discusses Health Care Initiatives," (May 1, 2006), <http://www.whitehouse.gov/news/releases/2006/05/20060501-5.html>.
7. Connecting California, California Regional Health Information Organization (CalRHIO), <http://calrhio.org/>.
8. Massachusetts eHealth Collaborative, <http://www.maehc.org/>.
9. The Delaware Health Information Network (DHIN), Delaware Health Care Commission (DHCC), <http://www.state.de.us/dhcc/information/dhin.shtml>.
10. HealthInfoNet, Maine Health Information Center, <http://www.hinfonet.org/>.
11. Canada Health Infoway, <http://www.infoway-inforoute.ca/en/home/home.aspx>.
12. National Health Information Network, U.S. Department of Health and Human Services, Office of the National Coordinator for Health Information Technology, <http://www.hhs.gov/healthit/nhin.html>.
13. NHS Connecting for Health Initiative, U.K. National Health Service (NHS), <http://www.connectingforhealth.nhs.uk/>.
14. S. Lohr, "High-Tech Alliance on Base for a Digital Health Network," *New York Times* (January 26, 2005), <http://www.nytimes.com/2005/01/26/technology/26health.html?ex=1160020800&en=62ff34612f19bd60&ei=5070>.
15. G. Gross, "IT Vendors Push for Open-Standards Health Network," IDG News Service (January 27, 2005), <http://www.computerworld.com.au/index.php/id/605768782;fp;16;fpid;0>.
16. NCQA's New Spine Care Recognition Program to Identify Doctors Who Deliver Patient-Centered, Evidence-Based Care, The National Committee for Quality Assurance (May 2, 2006), http://www.ncqa.org/Communications/News/SCRP_PublicComment.htm.
17. IBM Interoperable Health Information Infrastructure—IHII Project, IBM Research and Development and IBM Healthcare and Life Sciences, http://domino.research.ibm.com/comm/research_projects.nsf/pages/ihi.index.html.
18. *IT Infrastructure Technical Framework, Revision 2.0, Vol. 1 (ITI TF-1): Integration Profiles*, (August 15, 2005), Radiological Society of North America, Healthcare Information Management and Systems Society, and the American College of Cardiology, http://www.ihe.net/Technical_Framework/upload/ihe_iti_tf_2.0_vol1_FT_2005-08-15.pdf.
19. HSSP Implementation Subgroup (SOA for HL7), Healthcare Services Specification Project, <http://hssp-infrastructure.wikispaces.com/SOA+for+HL7>.
20. Resource Location and Updating Service (RLUS), Healthcare Services Specification Project, <http://hssp-rlus.wikispaces.com>.
21. Entity Identification Service (EIS), Healthcare Services Specification Project, <http://hssp-eis.wikispaces.com/>.
22. L. Haas, B. A. Eckman, P. Kodali, E. Lin, J. E. Rice, and P. M. Schwarz, "DiscoveryLink," in *Bioinformatics: Managing Scientific Data*, Z. Lacroix and T. Critchlow, Editors, Morgan Kaufmann Publishers, San Francisco, CA (2003), pp. 303–334.
23. R. Quinn, "Transaction Portal Cuts Costs. New York Payers and Providers Discover That IT Collaboration and the Sharing of Information Affords Savings That No Organization Could Achieve On Its Own," *Health Management Technology* **24**, No. 12, 40–42 (December 2003).
24. Standards and Specifications Overview, Utah Health Information Network (2006), <http://standards.uhin.com/overview.htm>.
25. Carolina Hospitals Start a RHIO, *Health Data Management* (January 26, 2006), <http://www.healthdatamanagement.com/portals/article.cfm?type=hospitals&articleid=12847>.
26. Western North Carolina Health Network to Link 16 Area Hospitals to Electronically Share Critical Patient Information, *MarketWire* (2006), http://www.marketwire.com/mw/release_html_b1?release_id=107528.
27. Health Records Put on Internet, *BBC News Online* (December 9, 2003), http://news.bbc.co.uk/2/hi/uk_news/england/3302705.stm.
28. Spine: NHS Care Records Service, National Health Service (United Kingdom), <http://www.connectingforhealth.nhs.uk/delivery/programmes/spine>.
29. J. H. Bjerregaard and P. C. Duedal, "MedCom: Danish Health Care Network," in *E-Health: Current Situation and Examples of Implemented and Beneficial E-Health Applications*, Volume 100, *Studies in Health Technology Informatics*, I. Iakkovidis, P. Wilson, and J. C. Healy, Editors, IOS Press, Amsterdam, The Netherlands, pp. 59–65 (September 2004).
30. The Danish eHealth Experience: One Portal for Citizens and Professionals, <http://www.dialog.sundhed.dk/NR/rdonlyres/ebkhhgtqfnuti6fyky74rwj7begja2grffb4cixasfccxmmopno6h3gdcgarwvavghb76lll4kwvhoazq2snzdgmee/The+Danish+eHealth+experience.pdf#search=The%20Danish%20eHealth%20experience%3A%20One%20Portal%20for%20Citizens%20and%20Professionals>.
31. R. Mack, S. Mukherjee, A. Soffer, N. Uramoto, E. Brown, A. Coden, J. Cooper, et al., "Text Analytics for Life Science Using the Unstructured Information Management

- Architecture," *IBM Systems Journal* **43**, No. 3, 490–515 (2004).
32. Healthcare and Life Sciences, IBM Corporation, <http://publib.boulder.ibm.com/infocenter/eserver/v1r2/index.jsp?topic=/ddqb/eicavhealthcareandlifesciences.htm>.
33. S. Renton, "The Bioterrorism Act of 2002, Protecting the Safety of Food and Drug Supply," *International Trade and Customs News* (September 2003).
34. M. A. Kuliasha, "Challenges of International Trade: Balancing Security and Commerce," <http://www.ornl.gov/~webworks/cpr/y2003/pres/116647.pdf>.
35. J. Heinrich, "SARS OUTBREAK: Improvements to Public Health Capacity Are Needed for Responding to Bioterrorism and Emerging Infectious Diseases," Testimony before the Subcommittee on Oversight and Investigations, House of Representatives, <http://www.gao.gov/new.items/d03769t.pdf>.
36. *Emerging Infections: Microbial Threats to Health in the United States*, J. Lederberg, R. E. Shope, and S. C. Oaks, Jr., Editors, The National Academies Press, Washington, D.C. (1992).
37. *Microbial Threats to Health: Emergence, Detection, and Response*, M. S. Smolinski, M. A. Hamburg, and J. Lederberg, Editors, The National Academies Press, Washington, D.C. (2003).
38. K. Green, "How Dangerous Is Anthrax? Risk and Policy Implications," Reason Public Policy Institute (2001), <http://www.rppi.org/wtc/anthrax.html>.
39. P. Wolfowitz, T. Ridge, and T. Thompson, *Deputy Secretary of Defense Paul Wolfowitz Joint Press Conference with the Department of Health and Human Services and Department of Homeland Security* (April 28, 2004), United States Department of Defense, <http://www.defenselink.mil/transcripts/2004/tr20040428-depsecdef1383.html>.
40. M. Wharton, T. L. Chorbha, R. L. Vogt, D. L. Morse, and J. W. Buehler, "Case Definitions for Public Health Surveillance," *Morbidity and Mortality Weekly Reports* **39**, No. RR-13, 1–43 (October 1990), <http://www.cdc.gov/mmwr/preview/mmwrhtml/00025629.htm>.
41. National Notifiable Diseases Surveillance System, Centers for Disease Control and Prevention, <http://www.cdc.gov/EPO/DPHSI/nndsshis.htm>.
42. Health Alert Network, Centers for Disease Control and Prevention, <http://www2a.cdc.gov/han/Index.asp>.
43. National Electronic Disease Surveillance System, Centers for Disease Control and Prevention, <http://www.cdc.gov/nedss/>.
44. The Epidemic Information Exchange, Centers for Disease Control and Prevention, <http://www.cdc.gov/mmwr/epix/epix.html>.
45. HL7 Reference Information Model (2006), Health Level Seven, Inc., www.hl7.org/library/data-model/RIM/modelpage_mem.htm.
46. PulseNet, Centers for Disease Control and Prevention, <http://www.cdc.gov/pulsenet/>.
47. Epidemic Intelligence Service, Centers for Disease Control and Prevention, <http://www.cdc.gov/eis/>.
48. *Emerging Infectious Diseases Journal*, Centers for Disease Control and Prevention, www.cdc.gov/ncidod/EID/index.htm.
49. Public Health Information Network (PHIN), Centers for Disease Control and Prevention, www.cdc.gov/phinf.
50. The Laboratory Response Network, Centers for Disease Control and Prevention, <http://www.bt.cdc.gov/lrn/>.
51. FSIS Establishes Food Emergency Response Network Division, U.S. Department of Agriculture, Food Safety and Inspection Service (FSIS), http://www.fsis.usda.gov/News_&_Events/NR_021505_01/index.asp.
52. The Electronic Laboratory Exchange Network, <https://www.elexnet.com/elex/index.jsp>.
53. The National Animal Health Laboratory Network (NAHLN), U. S. Department of Agriculture, Cooperative State Research, Education, and Extension Service, http://www.csrees.usda.gov/nea/ag_biosecurity/in_focus/apb_if_healthlab.html.
54. M. Enserink, "SARS in China: The Big Question Now: Will It Be Back?," *Science* **301**, No. 5631, 299 (2003).
55. M. Enserink, "SARS in China: China's Missed Chance," *Science* **301**, No. 5631, 294–296 (2003).
56. D. Normile and M. Enserink, "SARS in China" Tracking the Roots of a Killer," *Science* **301**, No. 5631, 297–299 (2003).
57. West Nile Virus, Division of Vector-Borne Infectious Diseases, Centers for Disease Control and Prevention, <http://www.cdc.gov/ncidod/dvbid/westnile/index.htm>.
58. Dengue Fever, Division of Vector-Borne Infectious Diseases, Centers for Disease Control and Prevention, <http://www.cdc.gov/ncidod/dvbid/dengue>.
59. Syndromic Surveillance: Putting Theory into Practice, *Fifth Annual Syndromic Surveillance Conference*, Baltimore, MD (2006), www.syndromic.org.
60. K. Hurt-Mullen, S. H. Lewis, C. Edwards, C. Jordan, R. A. Wojcik, and J. Lombardo, "Local Health Department Applications of Essence Biosurveillance System," *Proceedings of the 132nd Annual Meeting of the American Public Health Association*, Washington, D.C. (2004), http://apha.confex.com/apha/132am/techprogram/paper_93642.htm.
61. M. A. Hoffman, T. H. Wilkinson, A. Bush, W. Myers, R. G. Griffin, G. L. Hoff, and R. Archer, "Multijurisdictional Approach to Biosurveillance, Kansas City," *Emerging Infectious Diseases* **9**, No. 10, 1281–1286 (2003).
62. F. C. Tsui, J. U. Espino, V. M. Dato, P. H. Gesteland, J. Hutman, and M. M. Wagner, "Technical Description of RODS: A Real-Time Public Health Surveillance System," *Journal of the American Medical Informatics Association* **10**, No. 5, 399–408 (2003).
63. C.-S. Li, C. Aggrawal, M. Campbell, Y.-C. Chang, G. Glass, V. Iyengar, and M. Joshi, et al., "EPI-SPIRE: A Bio-Surveillance System for Environmental and Public Health Activity Monitoring," *Proceedings of the IEEE International Conference on Multimedia and Expo*, Baltimore, MD (2003), Volume 1, pp. 713–716.

Accepted for publication August 1, 2006.

Published online December 19, 2006.

Barbara A. Eckman

IBM Software Group, 1475 Phoenixville Pike, West Chester, Pennsylvania 19380 (baeckman@us.ibm.com). Dr. Eckman is a Senior Technical Staff Member in the Healthcare and Life Sciences Solution Development group, where she recently assumed the role of lead architect for Healthcare Interoperability. Her research interests in health care include the interface between clinical research and genomics and

systems biology research, and the extension of federated database technology to become a secure, efficient, fault-tolerant foundation for Regional Health Information Organizations. Other ongoing research interests center on bioinformatics data integration, where she has been active since the start of Human Genome Project funding in 1991. She has published over 20 academic papers and technical reports, and she has served on numerous organizing and program committees for bioinformatics research conferences as well as grant review panels for the National Institutes of Health and the National Science Foundation. Dr. Eckman received an M.S.E. degree in computer science from the University of Pennsylvania, an A.B. degree from Princeton University, and a Ph.D. degree in humanistic disciplines from the University of Pennsylvania.

Craig A. Bennett

IBM Application and Integration Middleware, 4111 Northside Parkway NW, Atlanta, Georgia 30327 (cbennett@us.ibm.com). Mr. Bennett is a Senior Managing Architect with IBM Global Business Transformation Outsourcing. He has a B.S. degree in computer engineering and an M.S. degree in computer science. His current responsibilities include defining solution architectures and providing technical and business process reengineering activities and delivery excellence for IBM clients in multiyear outsourcing contracts. Mr. Bennett has authored and coauthored numerous technical articles and holds several patents.

James H. Kaufman

IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120 (kaufman@almaden.ibm.com). Dr. Kaufman is manager of the Almaden Healthcare Research project. He received a B.A. degree from Cornell University and a Ph.D. degree from the University of California at Santa Barbara, both in physics. His current research includes interoperability for health care and public health, epidemiological modeling, and mathematical models of viral evolution. Dr. Kaufman is a Fellow of the American Physical Society.

Jeffrey W. Tenner

IBM Systems and Technology Group, 3605 HWY 52 N, Rochester, MN 55901 (tenner@us.ibm.com). Mr. Tenner is a Senior Technical Staff Member in the Healthcare and Life Sciences Development organization. He is also Chief Architect for the IBM/Mayo Clinic collaboration. He led the architecture and development of the IBM Clinical Genomics Solution and Data Discovery and Query Builder products. Previously, Mr. Tenner held technical leadership roles in the development of DB2® for iSeries™ and was lead architect for the Domino® portfolio of products for iSeries. Mr. Tenner holds several patents in database technology and is a member of the editorial board of the *IBM Systems Magazine*. ■